

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

In re Patent Application of  
Kwang-cheol OH et al  
Application No.: Unassigned  
Filing Date: November 26, 2003  
Title: VOICE REGION DETECTION APPARATUS AND METHOD

Group Art Unit: Unassigned  
Examiner: Unassigned  
Confirmation No.: Unassigned

**SUBMISSION OF CERTIFIED COPY OF PRIORITY DOCUMENT**

Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

Sir:

The benefit of the filing date of the following priority foreign application(s) in the following foreign country is hereby requested, and the right of priority provided in 35 U.S.C. § 119 is hereby claimed.

Country: Korea  
Patent Application No(s): 10-2002-0075650

Filed: November 30, 2002

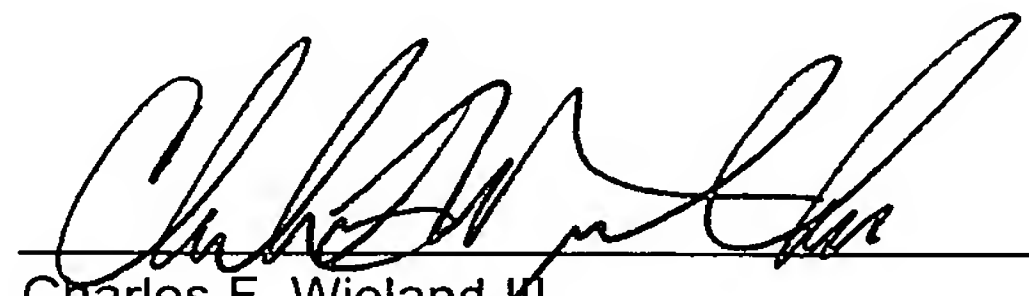
In support of this claim, enclosed is a certified copy(ies) of said foreign application(s). Said prior foreign application(s) is referred to in the oath or declaration. Acknowledgment of receipt of the certified copy(ies) is requested.

Respectfully submitted,

BURNS, DOANE, SWECKER & MATHIS, L.L.P.

P.O. Box 1404  
Alexandria, Virginia 22313-1404  
(703) 836-6620  
Date: November 26, 2003

By

  
Charles F. Wieland III  
Registration No. 33,096

(Translation)

**KOREAN INTELLECTUAL PROPERTY OFFICE**

This is to certify that the following application annexed hereto is a true copy from the records of the Korean Intellectual Property Office.

Application Number: 10-2002-0075650

Date of Application: November 30, 2002

Applicant(s): Samsung Electronics Co., Ltd.

Dated this 8th day of October, 2003

Commissioner (Seal)



별첨 사본은 아래 출원의 원본과 동일함을 증명함.

This is to certify that the following application annexed hereto is a true copy from the records of the Korean Intellectual Property Office.

출원 번호 : 10-2002-0075650  
Application Number

출원 년 월 일 : 2002년 11월 30일  
Date of Application NOV 30, 2002

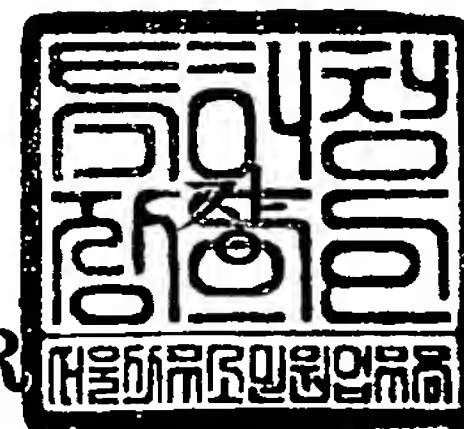
출원인 : 삼성전자주식회사  
Applicant(s) SAMSUNG ELECTRONICS CO., LTD.



2003 년 10 월 08 일

특 허 청

COMMISSIONER



## 【서지사항】

【서류명】	특허출원서
【권리구분】	특허
【수신처】	특허청장
【제출일자】	2002.11.30
【발명의 명칭】	음성구간 검출 장치 및 방법
【발명의 영문명칭】	APPARATUS AND METHOD OF VOICE REGION DETECTION
【출원인】	
【명칭】	삼성전자 주식회사
【출원인코드】	1-1998-104271-3
【대리인】	
【성명】	김동진
【대리인코드】	9-1999-000041-4
【포괄위임등록번호】	2002-007585-8
【발명자】	
【성명의 국문표기】	오광철
【성명의 영문표기】	OH,Kwang Cheol
【주민등록번호】	650413-1057814
【우편번호】	463-725
【주소】	경기도 성남시 분당구 금곡동 181번지 청솔한라아파트 303동 1204호
【국적】	KR
【발명자】	
【성명의 국문표기】	이영범
【성명의 영문표기】	LEE,Yong Beom
【주민등록번호】	611215-1024512
【우편번호】	442-744
【주소】	경기도 수원시 팔달구 영통동 황골마을벽산아파트 222동 1406호
【국적】	KR
【심사청구】	청구
【취지】	특허법 제42조의 규정에 의한 출원, 특허법 제60조의 규정에 의한 출원심사를 청구합니다. 대리인 김동진 (인)

1020020075650

출력 일자: 2003/10/15

【수수료】

【기본출원료】 20 면 29,000 원

【가산출원료】 16 면 16,000 원

【우선권주장료】 0 건 0 원

【심사청구료】 33 항 1,165,000 원

【합계】 1,210,000 원

【첨부서류】 1. 요약서·명세서(도면)\_1통

**【요약서】****【요약】**

본 발명은 유색잡음이 있는 음성신호에서도 음성구간을 정확하게 검출할 수 있는 음성구간 검출 장치 및 방법에 관한 것으로, 음성신호가 입력되면 입력된 음성신호를 프레임 단위로 나누어 프레임에 백색잡음을 합성하여 주변잡음을 백색화시킨 다음, 백색화된 프레임에서 프레임의 랜덤성을 나타내는 랜덤 파라미터를 추출하여 추출된 랜덤 파라미터값에 따라 프레임을 음성프레임과 잡음프레임으로 구분한 후, 이를 기초로 음성의 시작위치와 끝위치를 계산하여 음성구간을 검출함으로써, 다량의 유색잡음이 섞여 있는 음성신호에서도 정확하게 음성구간을 검출할 수 있도록 구성된 것을 특징으로 한다.

**【대표도】**

도 2

**【색인어】**

음성구간, 유색잡음, 랜덤, 파라미터

## 【명세서】

### 【발명의 명칭】

음성구간 검출 장치 및 방법{APPARATUS AND METHOD OF VOICE REGION DETECTION}

### 【도면의 간단한 설명】

도 1은 종래 음성구간 검출 장치의 동작을 설명하기 위한 도면이다.

도 2는 본 발명에 따른 음성구간 검출 장치의 개략적인 블록도이다.

도 3 내지 도 4는 프레임에서 주변잡음을 백색화하는 것을 설명하기 위한 도면이다.

도 5는 프레임에서 런의 갯수가 R일 확률  $P(R)$ 을 그래프로 나타낸 도면이다.

도 6은 프레임에서 랜덤 파라미터를 추출하는 것을 설명하기 위한 도면이다.

도 7은 본 발명에 따른 음성구간 검출 방법의 전체적인 흐름도이다.

도 8은 도 7에 있어서 프레임 상태 판단 단계의 상세 흐름도이다.

도 9는 프레임의 상태를 판단하는 방법을 설명하기 위한 도면이다.

도 10은 검출된 음성구간에서 유색잡음을 제거하는 방법을 설명하기 위한 도면이다.

도 11은 본 발명의 랜덤 파라미터에 따라 음성구간 검출 성능이 향상된 일예를 나타낸 도면이다.

\* 도면의 주요부분에 대한 부호의 설명 \*

10...전처리부

20...백색화부

21...백색잡음 발생부

22...신호 합성부

30...랜덤 파라미터 추출부

40...프레임 상태 판단부

50...음성구간 검출부

60...유색잡음 제거부

100...음성구간 검출 장치

【발명의 상세한 설명】

【발명의 목적】

【발명이 속하는 기술분야 및 그 분야의 종래기술】

- <17> 본 발명은 입력된 음성신호에서 음성구간을 검출하는 음성구간 검출 장치 및 방법에 관한 것으로, 특히 유색잡음이 있는 음성신호에서도 음성구간을 정확하게 검출할 수 있는 음성구간 검출 장치 및 방법에 관한 것이다.
- <18> 음성구간 검출은 외부로부터 입력된 음성신호에서 묵음 또는 잡음구간을 제외하고 순수한 음성구간만을 검출하기 위한 것으로, 그 대표적인 음성구간 검출 방법으로 음성신호의 에너지와 영교차율을 이용하여 음성구간을 검출하는 방법을 들 수 있다.
- <19> 그러나, 상기 음성구간 검출 방법은 주변잡음의 에너지가 큰 경우 무성음 구간과 같이 작은 에너지의 음성신호는 주변잡음에 묻혀버려 음성구간과 잡음구간을 구분하는 것이 매우 어렵다는 문제점이 있다.
- <20> 또한, 상기 음성구간 검출 방법은 마이크를 가까이 대고 음성을 입력하거나 임의로 마이크의 음량 레벨을 조절하면 음성신호의 입력 레벨이 달라지므로, 정확하게 음성구간을 검출하기 위해서는 입력장치 및 사용환경에 따라서 일일이 임계값을 수동으로 설정해야 하기 때문에 매우 번거롭다는 문제점이 있다.
- <21> 이러한 문제점을 해결하기 위한 것으로, 국내 공개특허 제2002-0030693호(발명의 명칭: 음성인식 시스템의 음성구간 결정 방법)에는 도 1(a)에 도시된 바와 같이 음성구간 검출시 음



성의 입력 레벨에 따라 임계값을 변화시켜줌으로써 주변잡음 및 입력장치에 관계없이 음성구간을 검출할 수 있는 방법이 개시되어 있다.

<22> 그러나, 상기 음성구간 결정 방법은 도 1(b)에 도시된 바와 같이 주변잡음이 백색잡음(white noise)인 경우에는 음성구간과 잡음구간을 명확하게 구별해낼 수 있지만, 도 1(c)에 도시된 바와 같이 주변잡음의 에너지가 크고 그 형태가 시간에 따라 변하는 유색잡음(color noise)인 경우 잡음구간과 음성구간이 잘 구별되지 않아 주변잡음을 음성구간으로 잘못 검출할 우려가 있다.

<23> 또한, 상기 음성구간 결정 방법은 반복적인 계산과정과 비교과정을 필요로 하기 때문에 이로 인하여 계산량이 많아져 실시간 사용이 어려울 뿐만 아니라, 마찰음의 스펙트럼 형태가 잡음과 유사하기 때문에 마찰음 구간을 정확하게 검출해낼 수 없어, 음성 인식의 경우와 같이 더욱 정확한 음성구간 검출이 요구되는 경우에는 부적합하다는 한계점이 있다.

#### 【발명이 이루고자 하는 기술적 과제】

<24> 본 발명은 상기한 문제점들을 해결하기 위해 안출된 것으로, 본 발명의 목적은 다량의 유색잡음이 섞여 있는 음성신호에서도 음성구간을 정확하게 검출할 수 있도록 하는 것이다.

<25> 본 발명의 다른 목적은 적은 계산량으로도 음성구간을 정확하게 검출하는 동시에 음성신호에서 주변잡음과 구별이 어려워 검출이 상대적으로 어려웠던 마찰음 구간도 검출할 수 있도록 하는 것이다.

#### 【발명의 구성 및 작용】

<26> 상기 목적을 달성하기 위하여 본 발명에 따른 음성구간 검출 장치는, 입력된 음성신호를 프레임 단위로 나누는 전처리부, 전처리부로부터 입력된 프레임에 백색잡음을 합성하는 백색

- 화부, 백색화부로부터 입력된 프레임에서 프레임의 랜덤성을 나타내는 랜덤 파라미터를 추출하는 랜덤 파라미터 추출부, 랜덤 파라미터 추출부를 통해 추출된 랜덤 파라미터값에 따라 프레임을 음성프레임과 잡음프레임으로 구분하는 프레임 상태 판단부, 및 프레임 상태 판단부로부터 입력된 음성프레임과 잡음프레임을 기초로 음성의 시작위치와 끝위치를 계산하여 음성구간을 검출하는 음성구간 검출부를 포함하는 것을 특징으로 한다.
- <27> 본 발명의 다른 바람직한 실시예에 있어서, 상기 음성구간 검출부를 통해 검출된 음성구간에서 유색잡음을 제거하는 유색잡음 제거부를 더 포함하는 것을 특징으로 한다.
- <28> 이하, 본 발명에 따른 음성구간 검출 장치의 구성과 동작에 대하여 첨부된 도면을 참조하여 상세히 설명한다.
- <29> 도 2는 본 발명에 따른 음성구간 검출 장치(100)의 개략적인 블록도로서, 도 2에 도시된 바와 같이 본 발명에 따른 음성구간 검출 장치(100)는 전처리부(10), 백색화부(20), 랜덤 파라미터 추출부(30), 프레임 상태 판단부(40), 음성구간 검출부(50), 및 유색잡음 제거부(60)를 포함한다.
- <30> 전처리부(10)는 입력된 음성신호를 소정 주파수로 샘플링한 후 샘플링된 음성신호를 음성 처리의 기본 단위인 프레임으로 나누는데, 본 발명에서는 8kHz로 샘플링된 음성에 대하여 160샘플(20ms) 단위로 하나의 프레임을 구성하였으며, 샘플링 비율 및 프레임당 샘플수는 적용 분야에 따라 변경이 가능하다.
- <31> 이렇게 해서 프레임 단위로 나뉘어진 음성신호는 백색화부(20)로 입력되는데, 백색화부(20)는 백색잡음 발생부(21)와 신호 합성부(22)를 통해 입력된 프레임에 백색잡음을 합성하여

주변잡음을 백색화(白色化, Whitening)시킴으로써 프레임내에서 주변잡음의 랜덤성을 증가시킨다.

<32> 백색잡음 발생부(21)는 주변잡음, 즉 비음성구간의 랜덤성을 강화하기 위하여 백색잡음을 발생시키는데, 백색잡음은 300Hz 내지 3500Hz와 같은 음성영역 내에서 그 기울기가 평탄한 주파수 스펙트럼을 가지는 균일 또는 가우시안 분포 신호로부터 생성되는 잡음이다. 여기에서, 백색잡음 발생부(21)에서 발생하는 백색잡음의 양은 주변잡음의 크기와 양에 따라 달라질 수 있는데, 본 발명에서는 음성신호의 초기 프레임들을 분석하여 백색잡음의 양을 설정하며, 이러한 설정과정은 음성구간 검출 장치(100)의 초기 구동시에 이루어질 수 있다.

<33> 신호 합성부(22)는 백색잡음 발생부(21)에서 발생한 백색잡음과 입력된 프레임 신호를 합성하기 위한 것으로, 일반적인 음성처리 분야에서 일반적으로 사용되는 신호 합성부와 그 구성 및 동작이 동일하므로 이에 대한 자세한 설명은 생략한다.

<34> 백색화부(20)를 통과한 프레임 신호의 일례가 도 3과 도 4에 도시되어 있는데, 도 3(a)는 입력된 음성신호, 도 3(b)는 도 3(a)의 음성신호에서 유성음 구간에 해당되는 프레임, 도 3(c)는 도 3(b)의 프레임에 백색잡음을 합성한 결과를 나타낸 도면이며, 도 4(a)는 입력된 음성신호, 도 4(b)는 도 4(a)의 음성신호에서 유색잡음구간에 해당되는 프레임, 도 4(c)는 도 4(b)의 프레임에 백색잡음을 합성한 결과를 나타낸 도면이다.

<35> 도 3에 도시된 바와 같이 유성음 구간에 해당되는 프레임 신호에 백색잡음을 합성하면 유성음 신호가 크기 때문에 영향을 거의 받지 않는 반면, 도 4에 도시된 바와 같이 잡음구간에 해당되는 프레임 신호에 백색잡음을 합성하면 잡음이 백색화되어 잡음 구간의 랜덤성이 증가되는 것을 알 수 있다.

- <36> 한편, 비교적 유색잡음이 없는 음성신호에서는 종래의 음성구간 검출 방법을 이용하여 만족할 만한 음성구간 검출 결과를 얻을 수 있지만, 주파수 스펙트럼의 분포가 일정하지 않은 유색잡음이 섞인 음성신호에서는 에너지나 영교차율 등의 파라미터로는 정확하게 잡음구간과 음성구간을 구분하기가 어렵다.
- <37> 따라서, 본 발명에서는 유색잡음이 섞인 음성신호에서도 음성구간을 정확하게 검출할 수 있도록 음성구간 판별을 위한 파라미터로 음성신호가 얼마나 랜덤한지를 나타내는 랜덤 파라미터를 이용하는데, 이하 랜덤 파라미터에 대하여 자세히 설명한다.
- <38> 본 발명에 있어서, 랜덤 파라미터란, 프레임의 랜덤성을 통계적 방식으로 테스트한 결과 값을 파라미터로 구성한 것을 의미하는데, 더 자세하게 설명하면, 비음성구간에서는 음성신호가 랜덤한 특성을 보이고 음성구간에서는 음성신호가 랜덤하지 않은 것을 이용하여, 확률 및 통계에서 사용되는 런 검증(run test)을 기반으로 프레임의 랜덤성을 수치로 나타낸 것이다.
- <39> 상기에서 런(run)은 연속된 시퀀스(sequence)에서 동일한 요소(elements)가 연속적으로 이어진 부시퀀스(sub-sequence), 즉, 같은 특성을 가지는 신호의 길이를 의미하는데, 예를 들면 시퀀스 「T H H H T H H T T T」에서 런의 수는 5개, 시퀀스 「S S S S S S S S S S S S R R R R R R R R R R」에서 런의 수는 2개, 시퀀스 「S R S R S R S R S R S R S R S R S R S R」에서 런의 수는 20개이며, 이러한 런의 갯수를 검증 통계량(test statistic)으로 하여 시퀀스의 랜덤성을 판단하는 것을 런 검증(run test)이라 한다.
- <40> 한편, 시퀀스내에서 런의 수가 너무 많아도 또는 너무 적어도 시퀀스는 랜덤하지 않은 것으로 판단되는데, 그 이유는 시퀀스 「S S S S S S S S S S S S R R R R R R R R R R R R」에서와 같이 시퀀스내에서 런의 갯수가 너무 작으면 "S" 또는 "R"이 연속적으로 위치하고 있을 확률이 높기 때문에 랜덤하지 않은 시퀀스로 판단되며, 시퀀스 「

<41> 따라서, 이러한 런 검증 개념을 프레임에 적용하여 프레임에서 런의 갯수를 검출하고 검출된 런의 갯수를 검증 통계량으로 하여 파라미터를 구성하면, 이 파라미터의 값에 따라 랜덤한 특성을 갖는 잡음구간과 주기적인 특성을 갖는 음성구간을 구별할 수 있는데, 본 발명에서 프레임의 랜덤성을 나타내는 랜덤 파라미터는 다음의 수학식 1과 같이 정의된다.

<43>      상기 수학적 식 1에 있어서, NR은 랜덤 파라미터(Number of Run), n은 프레임 길이의 1/2,  
R은 프레임내에서의 런의 갯수이다.

<45> 통계적 가설 검증(statistical hypothesis test)이란, 귀무가설(null hypothesis)/대립가설(alternative hypothesis)이 옳다는 전제하에서 검증 통계량(test statistic)의 값을 구한 후에 이 값이 나타날 가능성의 크기에 의하여 귀무가설/대립가설의 합리성 여부를 판단하는 가설 검증 방식으로, 이러한 통계적 가설 검증 방식에 따라 다음과 같이 "랜덤 파라미터는 프레임의 랜덤성을 나타내는 파라미터이다"라는 귀무가설을 검증한다.

37-10

는 가지수는  $\binom{n1+n2}{n1}$  이 되고,  $n1$ 개의 "0"중에서  $y1$ 개의 런을 발생시키는 가지수는  $\binom{n1-1}{y1-1}$  이 된다. 마찬가지로  $n2$ 개의 "1" 중에서  $y2$ 개의 런을 발생시키는 가지수는  $\binom{n2-1}{y2-1}$  이 된다. 따라서 하나의 프레임에서  $y1$ 개의 "0"런과  $y2$ 개의 "1"런이 발생할 확률은 다음의 식(1)과 같다.

<47>

$$P(y1,y2) = \frac{P(y1y2)}{P(y1)} = \frac{\binom{n1-1}{y1-1} \binom{n2-1}{y2-1}}{\binom{n1+n2}{n1}} \dots\dots\dots \text{식(1)}$$

<48> 한편, 프레임이 랜덤하다고 가정하면 프레임내에서 "0"과 "1"의 갯수는 거의 같다고 볼 수 있으며 "0"과 "1"에 대한 런의 갯수도 거의 같다고 볼 수 있다.

<49> 즉, 계산상의 편의를 위해,  $n1 \simeq n2 \simeq n$  ,  $y1 \simeq y2 \simeq y$  라 하면, 상기 식(1)은 다음의 식(2)와 같이 표현될 수 있다.

<50>

$$P(y,y) = \frac{\binom{n-1}{y-1} \binom{n-1}{y-1}}{\binom{2n}{n}} \dots\dots\dots \text{식(2)}$$

<51> 한편,  $n$ 개에서 임의로  $r$ 개를 뽑을 조합확률식  $nCr = \left[ \frac{n}{r} \right] = \frac{n!}{(n-r)!r!}$  에 따라 상기 식(2)를 정리하면, 상기 식(2)는 다음과 같은 과정을 통해 다음의 식(3)과 같이 표현될 수 있다.

<52>

$$P(y,y) = \frac{\frac{(n-1)!}{(n-y)!(y-1)!} \times \frac{(n-1)!}{(n-y)!(y-1)!}}{\frac{(2n)!}{n!n!}}$$

$$= \left( \frac{(n-1)!}{(n-y)!(y-1)!} \right)^2 \frac{n!n!}{(2n)!}$$

$$= \left( \frac{1}{(n-y)!(y-1)!} \right)^2 \frac{((n-1)!n!)^2}{(2n)!} \dots\dots\dots \text{식(3)}$$

<53> 따라서, 프레임내에 "0"에 대한 런의 갯수(y1)와 "1"에 대한 런의 갯수(y2)를 합쳐 총 R(R=y1+y2)개의 런이 있을 확률 P(R)은 다음 식(4)와 같이 표현될 수 있다.

<54>

$$P(R) \approx 2 \left( \frac{1}{(n-y)!(y-1)!} \right)^2 \frac{((n-1)!n!)^2}{(2n)!} \dots\dots\dots \text{식(4)}$$

<55> 상기 식(4)에서 알 수 있는 바와 같이, 프레임내에 총 R개의 런이 있을 확률 P(R)은 "0"과 1"에 대한 런의 갯수(y)를 변수로 하는 함수이므로, 따라서 런의 갯수(y)를 검증 통계량으로 설정할 수 있다.

<56> 도 5에 도시된 바와 같이, 프레임에서 런의 갯수가 R일 확률 P(R)을 그래프로 나타내면, 상기 확률 P(R)은 y=1 또는 y=n 일때 최소값, y=n/2일때 최대값을 가지며, 평균(E(R))과 분산(V(R))이 각각  $E(R)=n+1$ ,  $V(R)=\frac{n(n-1)}{2n-1}$  인 정규분포를 따르는 것을 알 수 있다.

<57> 한편, 정규분포를 따르는 확률 P(R)로 부터 에러율을 계산할 수 있으며, 이것은 도 5와 같은 정규분포에서의 확률은 곡선 아래 부분의 면적을 구하는 것과 같다. 즉, R의 평균(E(R))과 분산(V(R))으로부터 다음과 같은 식을 생각할 수 있다.

<58>

$$P(E(R) - \beta \sqrt{V(R)} < R < E(R) + \beta \sqrt{V(R)}) = \alpha \dots\dots\dots \text{식(5)}$$



- <59> 즉, 오차율은  $1-\alpha$ 로 나타나는데, 식(5)에서와 같이  $\beta$ 에 따라 조절할 수 있다. 즉,  $n$ 이 40일 때,  $\beta$ 가 1이면  $\alpha$ 는 0.6826이 되고,  $\beta$ 가 2이면  $\alpha$ 는 0.9544가 되고,  $\beta$ 가 3이면  $\alpha$ 는 0.9973이 된다. 즉 표준편차의 2배가 넘어가는 부분에 대해서 랜덤하지 않다고 결정하게 되면 4.56%의 에러를 포함하게 된다.
- <60> 따라서, "랜덤 파라미터는 프레임의 랜덤성을 나타내는 파라미터이다"라는 귀무가설을 기각할 수 없으므로, 랜덤 파라미터가 프레임의 랜덤성을 나타내는 파라미터인 것이 입증되었다.
- <61> 다시 도 2를 참조하면, 랜덤 파라미터 추출부(30)는 입력된 프레임에서 런의 갯수를 계산하여 계산에 의하여 얻어진 런의 갯수를 기초로 랜덤 파라미터를 추출하는데, 이하 도 6을 참조하여 프레임에서 랜덤 파라미터를 추출하는 방법에 대하여 설명한다.
- <62> 도 6은 프레임에서 랜덤 파라미터를 추출하는 방법을 설명하기 위한 도면으로, 도 6에 도시된 바와 같이 우선 입력된 프레임내의 샘플 데이터를 상위 비트쪽으로 1비트씩 쉬프트시키고 최하위 비트에는 0을 삽입한 후, 상기 1비트씩 쉬프트시켜 얻어진 프레임의 샘플 데이터와 원래 프레임의 샘플 데이터를 배타적 논리합 연산(exclusive OR operation)시킨다. 그 다음, 배타적 논리합 연산에 따른 결과값에서 "1"의 갯수, 즉, 프레임내에서의 런의 갯수를 계산한 후 이를 프레임 길이의 1/2로 나누어 이를 랜덤 파라미터로 추출한다.
- <63> 상기과 같은 과정을 거쳐 랜덤 파라미터 추출부(30)를 통해 랜덤 파라미터가 추출되면, 프레임 상태 판단부(40)는 추출된 랜덤 파라미터값에 따라 프레임의 상태를 판단하여 프레임을 음성성분을 가진 음성프레임과 잡음성분을 가진 잡음프레임으로 구분하는데, 추출된 랜덤 파라



- 미터값에 따라 프레임의 상태를 판단하는 방법에 대하여는 도 8에 관한 설명에서 자세히 서술하기로 한다.

<64> · 음성구간 검출부(50)는 프레임 상태 판단부(40)로부터 입력된 음성프레임과 잡음프레임을 기초로 음성의 시작위치와 끝위치를 계산하여 음성구간을 검출한다.

<65> 한편, 입력된 음성신호에 다량의 유색잡음이 섞여 있는 경우, 음성구간 검출부(50)를 통해 검출된 음성구간에는 일부 유색잡음이 포함될 수도 있는데, 이를 위하여, 본 발명에서는 음성구간 검출부(50)에서 검출된 음성구간에 유색잡음이 섞여 있다고 판단되면, 유색잡음 제거부(60)를 통해 유색잡음의 특성을 찾아내서 이를 제거한 후 유색잡음이 제거된 음성구간을 다시 랜덤 파라미터 추출부(30)로 출력한다.

<66> 여기에서, 잡음 제거 방법으로는 간단하게 주변잡음으로 추정되는 구간에서 LPC계수를 구하고 음성구간에 대해 전체적으로 LPC 역필터링하는 방법을 사용할 수 있다.

<67> 유색잡음이 제거된 음성구간의 프레임들이 랜덤 파라미터 추출부(30)로 입력되면, 다시 상기와 같은 랜덤 파라미터 추출, 프레임 상태 판단, 음성구간 검출 과정을 거치게 되며, 이로 인하여 음성구간에 유색잡음이 포함될 가능성을 최소화시킬 수 있다.

<68> 따라서, 유색잡음 제거부(60)를 통해 음성구간에 섞여 있는 유색잡음을 제거함으로써, 다량의 유색잡음이 섞여 있는 음성신호가 입력되어도 정확하게 음성구간만을 검출할 수 있다.

<69> 한편, 본 발명에 따른 음성구간 검출 방법은, 음성신호가 입력되면 입력된 음성신호를 프레임으로 나누는 단계, 프레임에 백색잡음을 합성하여 주변잡음을 백색화시키는 단계, 백색화된 프레임에서 프레임의 랜덤성을 나타내는 랜덤 파라미터를 추출하는 단계, 추출된 랜덤 파라미터값에 따라 프레임을 음성프레임과 잡음프레임으로 구분하는 단계, 및 복수개의 음성프레

임과 잡음프레임을 기초로 음성의 시작위치와 끝위치를 계산하여 음성구간을 검출하는 단계를 포함하는 것을 특징으로 한다.

<70> 이하, 본 발명에 따른 음성 검출 방법에 대하여 첨부된 도면들을 참조하여 상세히 설명한다.

<71> 도 7은 본 발명에 따른 음성 검출 방법의 흐름도이다.

<72> 우선, 음성신호가 입력되면 전처리부(10)를 통해 입력된 음성신호를 소정 주파수로 샘플링한 후 샘플링된 음성신호를 음성 처리의 기본 단위인 프레임으로 나눈다(S10).

<73> 여기에서, 프레임 사이의 간격은 가급적 작게 하여 음소성분을 정확히 파악할 수 있도록 하고, 프레임은 서로 중복시켜 프레임 사이에서 데이터 손실을 방지할 수 있도록 하는 것이 바람직하다.

<74> 그 다음, 백색화부(20)는 입력된 프레임에 백색잡음을 합성하여 주변잡음을 백색화시키는데(S20), 프레임에 백색잡음을 합성하면 프레임에 섞여 있는 잡음성분의 랜덤성이 증가되어 음성구간 검출시 랜덤한 특성을 갖는 잡음구간과 주기적인 특성을 갖는 음성구간이 확실하게 구별되기 때문이다.

<75> 그 다음, 랜덤 파라미터 추출부(30)는 프레임에서 런의 갯수를 계산하여 계산에 의하여 얻어진 런의 갯수를 기초로 랜덤 파라미터를 추출하는데(S30), 랜덤 파라미터를 추출하는 방법에 대하여는 도 6과 관련된 설명에서 상세히 설명하였으므로 이에 대한 자세한 설명은 생략한다.

<76> 그 다음, 프레임 상태 판단부(40)는 랜덤 파라미터 추출부(30)에서 추출된 랜덤 파라미터값에 따라 프레임의 상태를 판단하여 프레임을 음성프레임과 잡음프레임으로 구분하는데

- (S40), 이하 도 8 및 도 9를 참조하여 프레임 상태 판단 단계(S40)에 대하여 더 자세히 설명한다.

<77> 도 8은 도 7에 있어서 프레임 상태 판단 단계(S40)의 상세 흐름도이며, 도 9는 프레임 상태 판단을 위한 임계값 설정을 설명하기 위한 도면이다.

<78> 여러 프레임들에서 랜덤 파라미터를 추출해본 결과, 랜덤 파라미터는 0에서 2사이의 값을 가지는데, 특히 랜덤한 특성을 가지는 잡음 구간에서는 1에 가까운 값을, 유성음을 포함한 일반적인 음성구간에서는 0.8 이하의 값을, 마찰음 구간에서는 1.2 이상의 값을 갖는 특성이 있다.

<79> 따라서, 본 발명에서는 이러한 랜덤 파라미터의 특성을 이용하여 도 9에 도시된 바와 같이 추출된 랜덤 파라미터값에 따라 프레임의 상태를 판단하여 프레임을 음성성분을 가진 음성 프레임과 잡음성분을 가진 잡음프레임으로 구분하는데, 특히, 유성음인지 마찰음인지를 판단할 수 있는 기준값을 각각 제1 임계값, 제2 임계값으로 미리 설정해 놓고, 프레임의 랜덤 파라미터값을 상기 제1,2 임계값과 비교함으로써, 음성프레임에서도 유성음 프레임과 마찰음 프레임을 각각 구분할 수 있도록 하였다. 상기에서, 제1 임계값은 0.8, 제2 임계값은 1.2인 것이 바람직하다.

<80> 즉, 프레임 상태 판단부(40)는 랜덤 파라미터값이 제1 임계값 이하이면 해당 프레임을 유성음 프레임으로 판단하고(S41~S42), 랜덤 파라미터값이 제2 임계값 이상이면 해당 프레임을 마찰음 프레임으로 판단하며(S43~S44), 랜덤 파라미터값이 제1 임계값 이상 제2 임계값 이하이면 해당 프레임을 잡음프레임으로 판단한다(S45).

- <81>        그 다음, 입력된 음성신호의 모든 프레임에 대해 프레임 상태 판단이 완료되었는지를 체크하여(S50), 모든 프레임에 대해 프레임 상태 판단이 완료되었으면 프레임 상태를 통해 검출된 복수개의 유성음 프레임, 마찰음 프레임, 잡음 프레임을 기초로 음성의 시작위치와 끝 위치를 계산하여 음성구간을 검출하며(S60), 그렇지 않은 경우에는 다음 프레임에 대해 상기와 같은 백색화, 랜덤 파라미터 추출, 프레임 상태 판단 과정을 수행한다.
- <82>        한편, 입력된 음성신호에 다량의 유색잡음이 섞여 있는 경우, 상기 음성구간 검출 단계(S60)를 통해 검출된 음성구간에 일부 유색잡음이 포함될 가능성이 있다.
- <83>        따라서, 본 발명에서는 음성구간 검출의 신뢰성을 향상시키기 위하여 검출된 음성구간에 유색잡음이 섞여 있다고 판단되면 음성구간에 포함된 유색잡음의 특성을 찾아내서 제거하는데(S70~S80), 이하 도 10을 참조하여 유색잡음 제거 단계(S70~S80)에 대하여 더 자세히 설명한다.
- <84>        도 10은 검출된 음성구간에서 유색잡음을 제거하는 방법을 설명하기 위한 도면으로, 도 10(a)는 유색잡음이 섞여 있는 음성신호, 도 10(b)는 도 10(a)의 음성신호에 대한 랜덤 파라미터, 도 10(c)는 도 10(a)의 음성신호에서 유색잡음을 제거한 후 랜덤 파라미터를 추출한 결과를 나타낸 도면이다.
- <85>        도 10(b)에 도시된 바와 같이 유색잡음이 섞여 있는 음성신호에서 랜덤 파라미터를 추출해 보면, 유색잡음으로 인하여 랜덤 파라미터값이 도 10(c)와 비교하여 전체적으로 0.1 내지 0.2 정도 낮은 것을 알 수 있으며, 따라서 이러한 랜덤 파라미터의 특성을 이용하면 음성구간 검출부(50)를 통해 검출된 음성구간에 유색잡음이 섞여 있는지의 여부를 판단할 수 있다.

- <86> 도 9에 도시된 바와 같이, 유색잡음으로 인한 랜덤 파라미터의 감소량을  $\Delta d$  라 하면, 검출된 음성구간의 랜덤 파라미터 평균값이 제1 임계값을 기준으로  $\Delta d$  이하이거나, 검출된 음성구간의 랜덤 파라미터 평균값이 제2 임계값을 기준으로  $\Delta d$  이하인 경우 음성구간에 유색잡음이 섞여 있는 것으로 판단할 수 있다.
- <87> 즉, 유색잡음 제거부(60)는 음성구간 검출부(50)를 통해 검출된 음성구간에서 랜덤 파라미터들의 평균값을 계산하여 계산된 랜덤 파라미터 평균값이 제1 임계값- $\Delta d$  이하이거나, 계산된 랜덤 파라미터 평균값이 제2 임계값- $\Delta d$  이하이면, 검출된 음성구간에 유색잡음이 섞여 있다고 판단한다.
- <88> 상기에서, 제1 임계값은 0.8, 제2 임계값은 1.2인 것이 바람직하며, 유색잡음으로 인한 랜덤 파라미터의 감소량  $\Delta d$ 는 0.1 내지 0.2인 것이 바람직하다.
- <89> 그 다음, 이러한 과정을 거쳐 음성구간에 유색잡음이 섞여 있다고 판단되면 유색잡음 제거부(60)는 음성구간에 포함된 유색잡음의 특성을 찾아내서 제거하는데(S80), 잡음 제거 방법으로는 간단하게 주변잡음으로 추정되는 구간에서 LPC계수를 구하고 음성구간에 대해 전체적으로 LPC 역필터링하는 방법을 사용할 수 있으며, 이 외에 다른 잡음 제거 방법을 사용하는 것도 가능하다.
- <90> 그 다음, 유색잡음이 제거된 음성구간의 프레임들은 다시 랜덤 파라미터 추출부(30)로 입력되어 다시 상기와 같은 랜덤 파라미터 추출, 프레임 상태 판단, 음성구간 검출 과정을 거치게 되며, 이로 인하여 음성구간에 유색잡음이 포함될 가능성을 최소화시킬 수 있으므로 유색잡음이 섞여 있는 음성신호에서 음성구간만을 정확하게 검출할 수 있다.

- <91> 도 11은 본 발명의 랜덤 파라미터에 따라 음성구간 검출 성능이 향상된 일예를 나타낸 도면으로, 도 11(a)는 핸드폰 단말기에서 녹음된 음성신호 "스프레트쉬트"를 나타낸 도면이고, 도 11(b)는 도 11(a)의 음성신호에 대한 평균 에너지를 나타낸 도면이며, 도 11(c)는 도 11(a)의 음성신호에 대한 랜덤 파라미터를 나타낸 도면이다.
- <92> 도 11(b)에 도시된 바와 같이 종래의 에너지 파라미터를 이용하면 유색잡음에 의하여 음성신호에서 "스프"에 대한 구간이 마스킹되어 음성구간 검출이 제대로 이루어질 수 없는 반면, 도 11(c)에 도시된 바와 같이 본 발명의 랜덤 파라미터를 이용하면 유색잡음이 섞여 있는 음성신호에서도 음성구간과 잡음구간을 확실하게 구별해낼 수 있다.
- <93> 본 발명은 도면에 도시된 일 실시예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다. 따라서, 본 발명의 진정한 기술적 보호범위는 첨부된 청구범위의 기술적 사상에 의해 정해져야 할 것이다.

#### 【발명의 효과】

- <94> 상기한 바와 같이, 본 발명의 음성구간 검출 장치 및 방법에 따르면, 다량의 유색잡음이 섞여 있는 음성신호에서도 정확하게 음성구간을 검출할 수 있을 뿐만 아니라, 잡음과 구별이 어려워 검출이 상대적으로 어려웠던 마찰음도 정확하게 검출할 수 있으므로, 정확한 음성구간 검출을 필요로 하는 음성 인식, 화자 인식 시스템의 성능을 향상시킬 수 있는 효과가 있다.
- <95> 또한, 본 발명에 따르면 음성구간 검출을 위한 임계값을 환경에 따라 변화시키지 않고도 정확하게 음성구간을 검출할 수 있어 불필요한 계산량을 줄일 수 있는 효과도 있다.

<96> 또한, 본 발명에 따르면 무음 구간 및 잡음 구간을 음성신호로 간주하여 처리하는데 따른 메모리 용량의 증대방지가 가능하며, 음성구간만을 추출하여 처리함으로써 처리시간의 단축이 가능하게 된다.

【특허청구범위】

【청구항 1】

입력된 음성신호를 프레임 단위로 나누는 전처리부;

상기 전처리부로부터 입력된 프레임에 백색잡음을 합성하는 백색화부;

상기 백색화부로부터 입력된 프레임에서 프레임의 랜덤성을 나타내는 랜덤 파라미터를 추출하는 랜덤 파라미터 추출부;

상기 랜덤 파라미터 추출부를 통해 추출된 랜덤 파라미터값에 따라 프레임을 음성프레임과 잡음프레임으로 구분하는 프레임 상태 판단부; 및

상기 프레임 상태 판단부로부터 입력된 음성프레임과 잡음프레임을 기초로 음성의 시작 위치와 끝위치를 계산하여 음성구간을 검출하는 음성구간 검출부를 포함하는 것을 특징으로 하는 음성구간 검출 장치.

【청구항 2】

제 1항에 있어서, 상기 전처리부는,

상기 입력된 음성신호를 소정 주파수로 샘플링한 후 샘플링된 음성신호를 다수의 프레임으로 나누는 것을 특징으로 하는 음성구간 검출 장치.

【청구항 3】

제 2항에 있어서, 상기 다수의 프레임은 서로 중복되는 것을 특징으로 하는 음성구간 검출 장치.

【청구항 4】

제 1항에 있어서, 상기 백색화부는,



백색잡음을 발생시키는 백색잡음 발생부; 및

상기 백색잡음 발생부에서 발생된 백색잡음과 상기 전처리부로부터 입력된 프레임 신호를 합성하는 신호 합성부를 포함하는 것을 특징으로 하는 음성구간 검출 장치.

#### 【청구항 5】

제 1항, 제 2항, 제 3항 또는 제 4항에 있어서, 상기 랜덤 파라미터 추출부는, 상기 백색화부를 통해 백색화된 프레임에서 동일한 요소가 연속적으로 이어진 런의 갯수를 계산한 후 상기 계산된 런의 갯수를 기초로 랜덤 파라미터를 추출하는 것을 특징으로 하는 음성구간 검출 장치.

#### 【청구항 6】

제 5항에 있어서, 상기 랜덤 파라미터는,

$$NR = \frac{R}{n}$$

(단, NR은 랜덤 파라미터, n은 프레임 길이의 1/2, R은 프레임내에서의 런의 갯수)

인 것을 특징으로 하는 음성구간 검출 장치.

#### 【청구항 7】

제 1항 또는 제 6항에 있어서, 상기 음성프레임은 유성음 프레임과 마찰음 프레임을 포함하는 것을 특징으로 하는 음성구간 검출 장치.

【청구항 8】

제 7항에 있어서, 상기 프레임 상태 판단부는,

상기 랜덤 파라미터 추출부에서 추출된 랜덤 파라미터값이 제1 임계값 이하이면 해당 프레임을 유성음 프레임으로 판단하는 것을 특징으로 하는 음성구간 검출 장치.

【청구항 9】

제 8항에 있어서, 상기 제1 임계값은 0.8인 것을 특징으로 하는 음성구간 검출 장치.

【청구항 10】

제 8항에 있어서, 상기 프레임 상태 판단부는,

상기 랜덤 파라미터 추출부에서 추출된 랜덤 파라미터값이 제2 임계값 이상이면 해당 프레임을 마찰음 프레임으로 판단하는 것을 특징으로 하는 음성구간 검출 장치.

【청구항 11】

제 10항에 있어서, 상기 제2 임계값은 1.2인 것을 특징으로 하는 음성구간 검출 장치.

【청구항 12】

제 10항에 있어서, 상기 프레임 상태 판단부는,

상기 랜덤 파라미터 추출부에서 추출된 랜덤 파라미터값이 상기 제1 임계값 이상이고 상기 제2 임계값 이하이면, 해당 프레임을 잡음프레임으로 판단하는 것을 특징으로 하는 음성구간 검출 장치.

【청구항 13】

제 12항에 있어서, 상기 제1 임계값은 0.8이고, 상기 제2 임계값은 1.2인 것을 특징으로 하는 음성구간 검출 장치

【청구항 14】

제 1항에 있어서, 상기 음성구간 검출부를 통해 검출된 음성구간에서 유색잡음을 제거하는 유색잡음 제거부를 더 포함하는 것을 특징으로 하는 음성구간 검출 장치.

【청구항 15】

제 10항에 있어서, 상기 음성구간 검출부를 통해 검출된 음성구간에서 유색잡음을 제거하는 유색잡음 제거부를 더 포함하되,

상기 유색잡음 제거부는 상기 음성구간 검출부를 통해 검출된 음성구간의 랜덤 파라미터 평균값이 소정 임계값 이하일 경우 상기 검출된 음성구간에서 유색잡음을 제거하는 것을 특징으로 하는 음성구간 검출 장치.

【청구항 16】

제 15항에 있어서, 상기 소정 임계값은 상기 제1 임계값에서 유색잡음에 의한 랜덤 파라미터의 감소량을 뺀 값인 것을 특징으로 하는 음성구간 검출 장치.

【청구항 17】

제 15항에 있어서, 상기 소정 임계값은 상기 제2 임계값에서 유색잡음에 의한 랜덤 파라미터의 감소량을 뺀 값인 것을 특징으로 하는 음성구간 검출 장치.

【청구항 18】

음성신호가 입력되면 입력된 음성신호를 프레임으로 나누는 단계;

상기 프레임에 백색잡음을 합성하여 주변잡음을 백색화시키는 단계;

상기 백색화된 프레임에서 프레임의 랜덤성을 나타내는 랜덤 파라미터를 추출하는 단계;

상기 추출된 랜덤 파라미터값에 따라 프레임을 음성프레임과 잡음프레임으로 구분하는 단계; 및

상기 음성프레임과 잡음프레임을 기초로 음성의 시작위치와 끝위치를 계산하여 음성구간을 검출하는 단계를 포함하는 것을 특징으로 하는 음성구간 검출 방법.

【청구항 19】

제 18항에 있어서, 상기 입력된 음성신호를 프레임으로 나누는 단계는,

상기 입력된 음성신호를 소정 주파수로 샘플링한 후 샘플링된 음성신호를 다수의 프레임으로 나누는 단계를 포함하는 것을 특징으로 하는 음성구간 검출 방법.

【청구항 20】

제 19항에 있어서, 상기 다수의 프레임은 서로 중복되는 것을 특징으로 하는 음성구간 검출 방법.

【청구항 21】

제 18항에 있어서, 상기 주변잡음을 백색화시키는 단계는,

백색잡음을 발생시키는 단계; 및

상기 발생된 백색잡음과 상기 프레임 신호를 합성하는 단계를 포함하는 것을 특징으로 하는 음성구간 검출 방법.

## 【청구항 22】

제 18항, 제 19항, 제 20항 또는 제 21항에 있어서, 상기 랜덤 파라미터를 추출하는 단계는,

상기 백색화된 프레임에서 동일한 요소가 연속적으로 이어진 런의 갯수를 계산하는 단계; 및

상기 계산된 런의 갯수를 프레임 길이로 나누어 이를 랜덤 파라미터로 추출하는 단계를 포함하는 것을 특징으로 하는 음성구간 검출 방법.

## 【청구항 23】

제 22항에 있어서, 상기 랜덤 파라미터는,

$$NR = \frac{R}{n}$$

(단, NR은 랜덤 파라미터, n은 프레임 길이의 1/2, R은 프레임내에서의 런의 갯수)

인 것을 특징으로 하는 음성구간 검출 장치.

## 【청구항 24】

제 18항 또는 제 23항에 있어서, 상기 음성프레임은 유성음 프레임과 마찰음 프레임을 포함하는 것을 특징으로 하는 음성구간 검출 방법.

## 【청구항 25】

제 24항에 있어서, 상기 추출된 랜덤 파라미터값이 제1 임계값 이하이면 해당 프레임을 유성음 프레임으로 판단하는 단계를 포함하는 것을 특징으로 하는 음성구간 검출 방법.

## 【청구항 26】

제 25항에 있어서, 상기 제1 임계값은 0.8인 것을 특징으로 하는 음성구간 검출 방법.

## 【청구항 27】

제 25항에 있어서, 상기 추출된 랜덤 파라미터값이 제2 임계값 이상이면 해당 프레임을 마찰음 프레임으로 판단하는 단계를 포함하는 것을 특징으로 하는 음성구간 검출 방법.

## 【청구항 28】

제 27항에 있어서, 상기 제2 임계값은 1.2인 것을 특징으로 하는 음성구간 검출 방법.

## 【청구항 29】

제 27항에 있어서, 상기 추출된 랜덤 파라미터값이 상기 제1 임계값 이상이고 상기 제2 임계값 이하이면, 해당 프레임을 잡음프레임으로 판단하는 단계를 포함하는 것을 특징으로 하는 음성구간 검출 방법.

## 【청구항 30】

제 29항에 있어서, 상기 제1 임계값은 0.8이고, 상기 제2 임계값은 1.2인 것을 특징으로 하는 음성구간 검출 장치.

## 【청구항 31】

제 27항에 있어서, 상기 검출된 음성구간의 랜덤 파라미터 평균값이 소정 임계값 이하일 경우 상기 검출된 음성구간에서 유색잡음을 제거하는 단계를 더 포함하는 것을 특징으로 하는 음성구간 검출 방법.

【청구항 32】

제 31항에 있어서, 상기 소정 임계값은 상기 제1 임계값에서 유색잡음에 의한 랜덤 파라미터의 감소량을 뺀 값인 것을 특징으로 하는 음성구간 검출 방법.

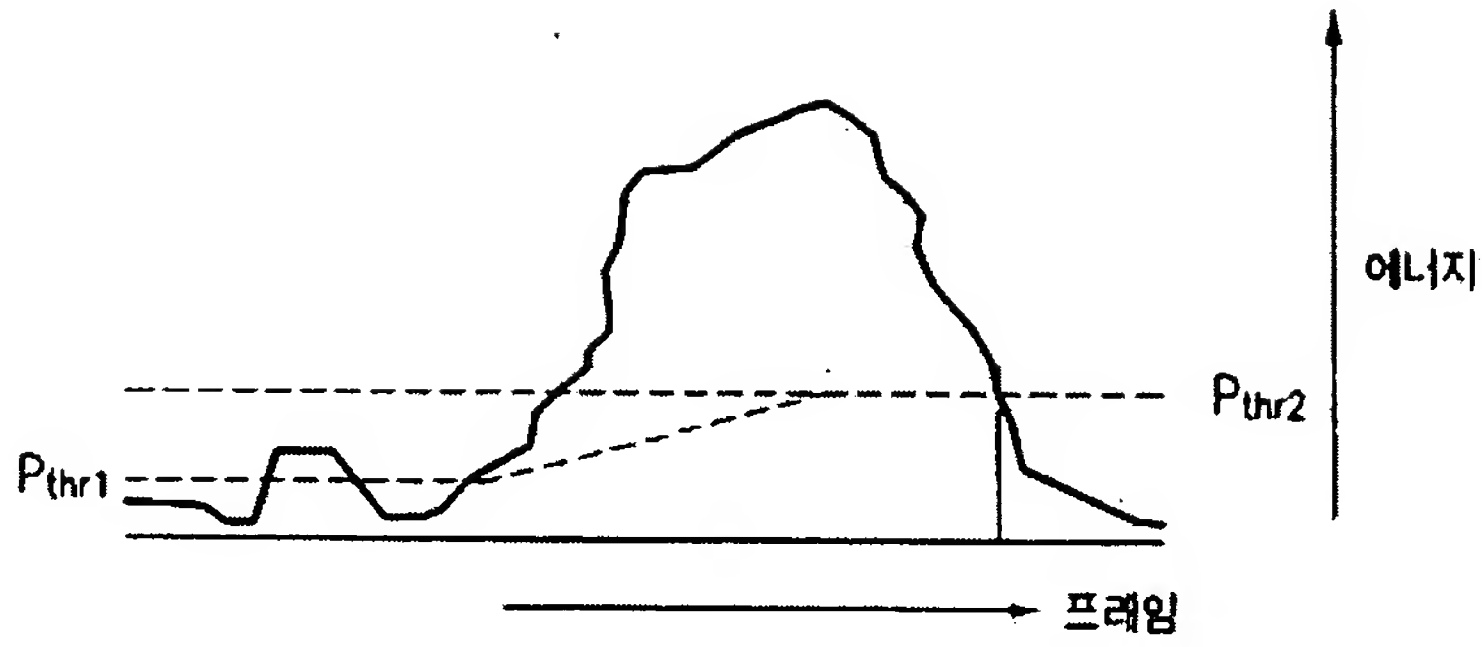
【청구항 33】

제 31항에 있어서, 상기 소정 임계값은 상기 제2 임계값에서 유색잡음에 의한 랜덤 파라미터의 감소량을 뺀 값인 것을 특징으로 하는 음성구간 검출 방법.

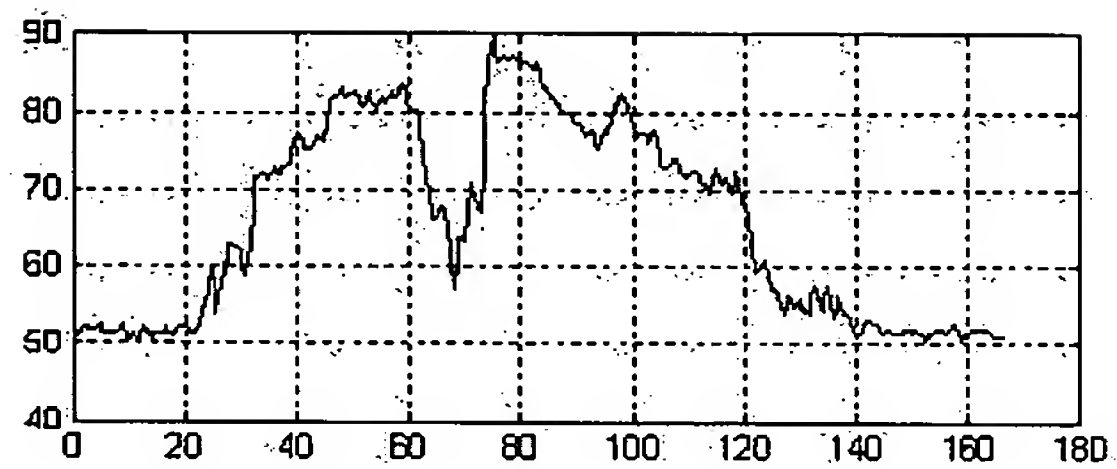
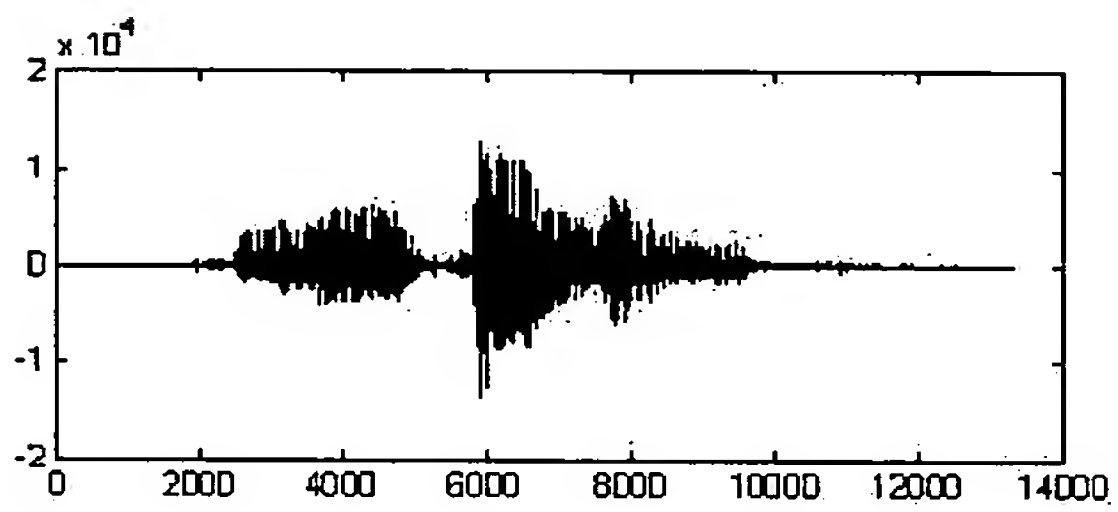
【도면】

【도 1】

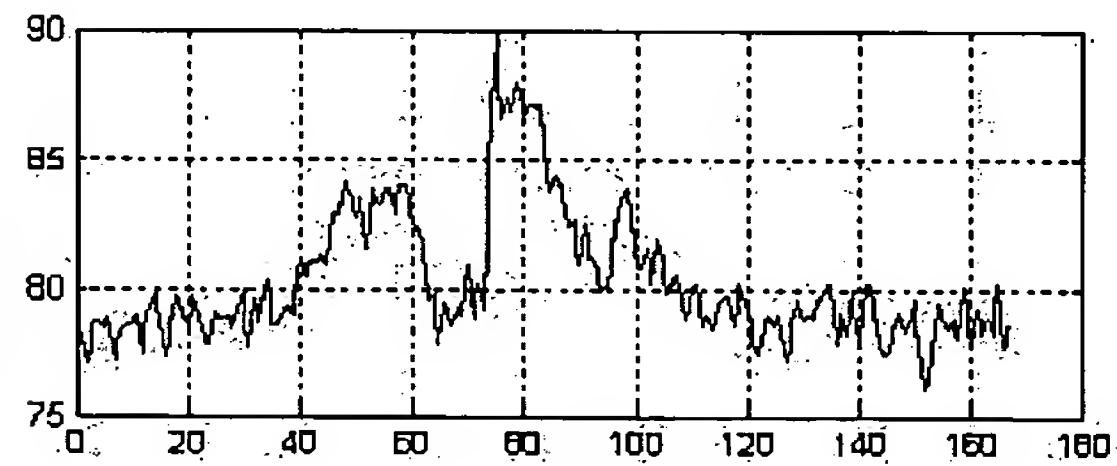
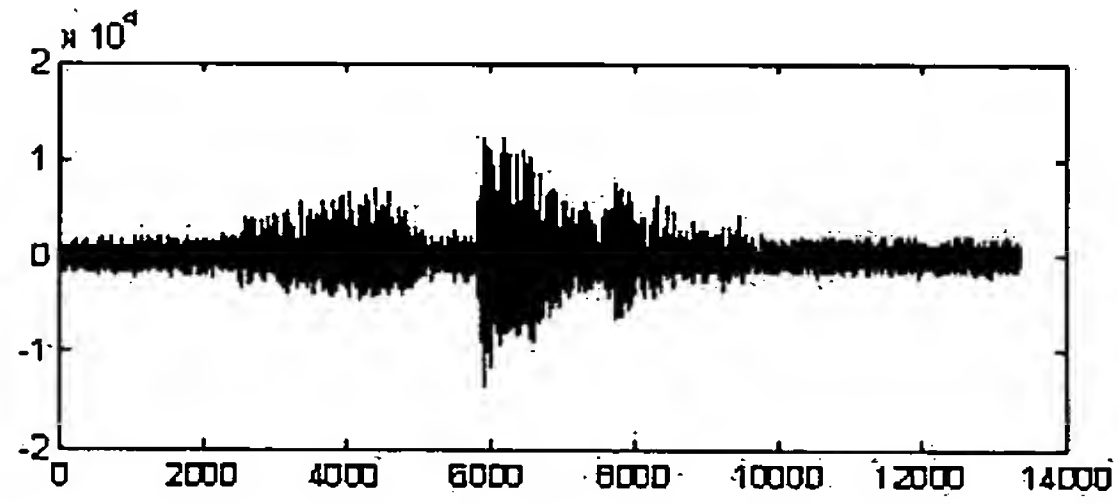
(a)



(b)

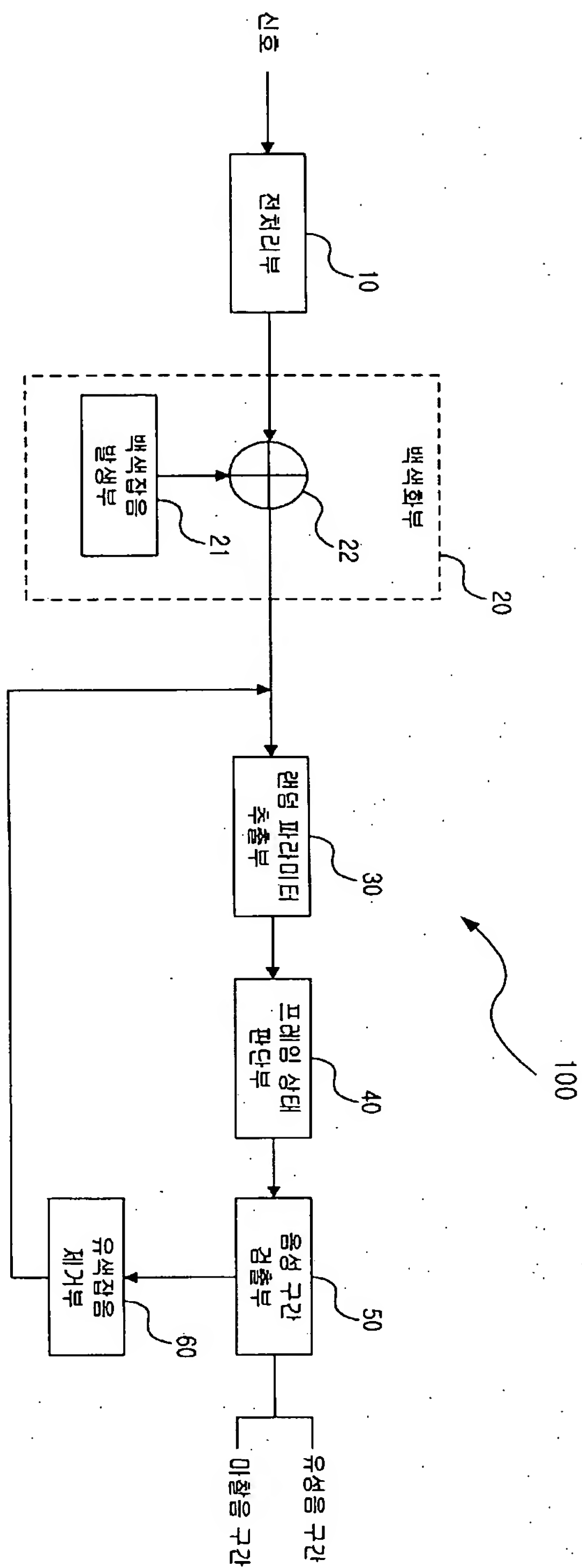


(c)

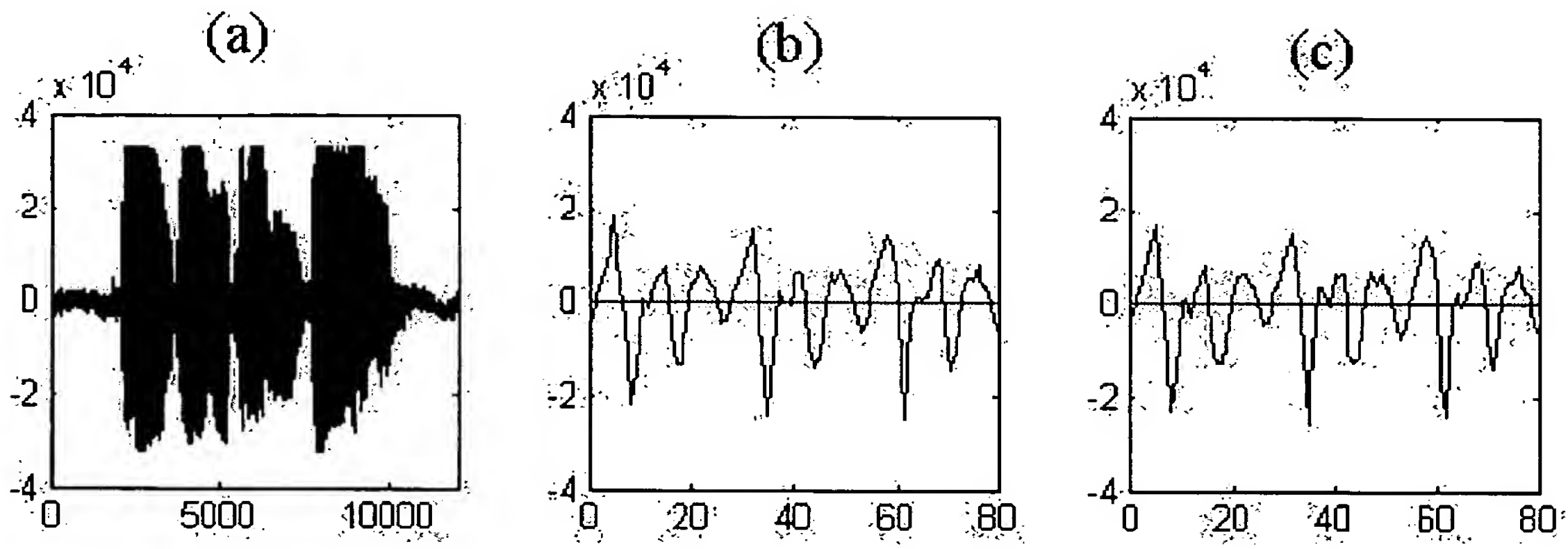




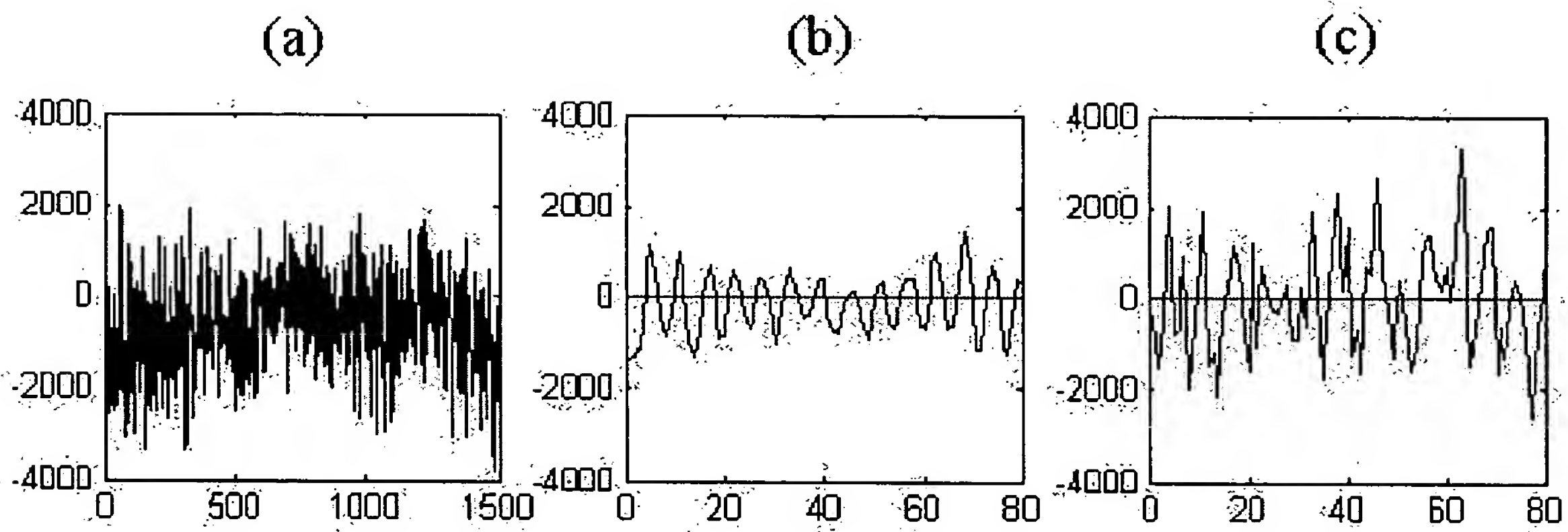
【도, 2】



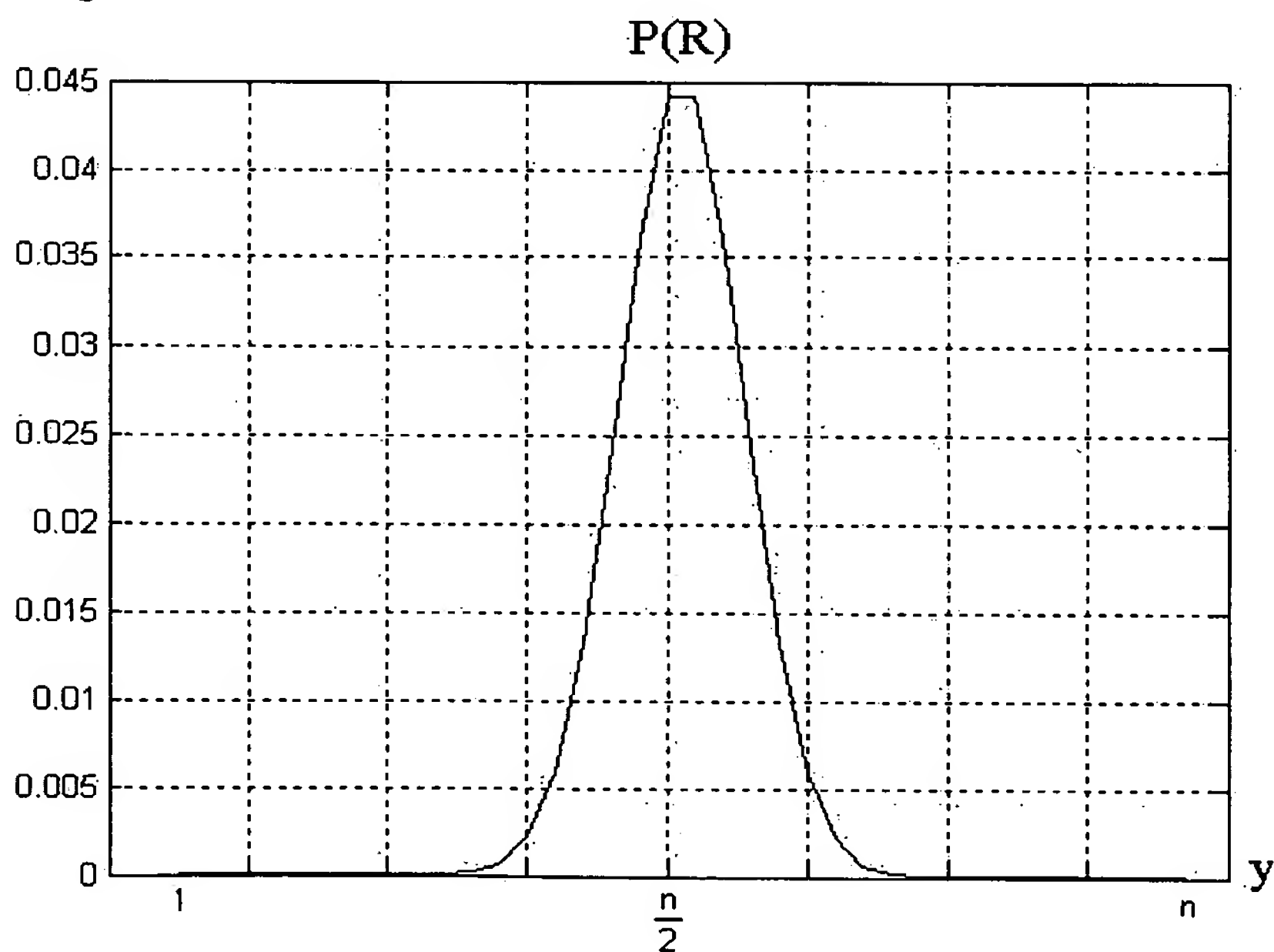
【도 3】



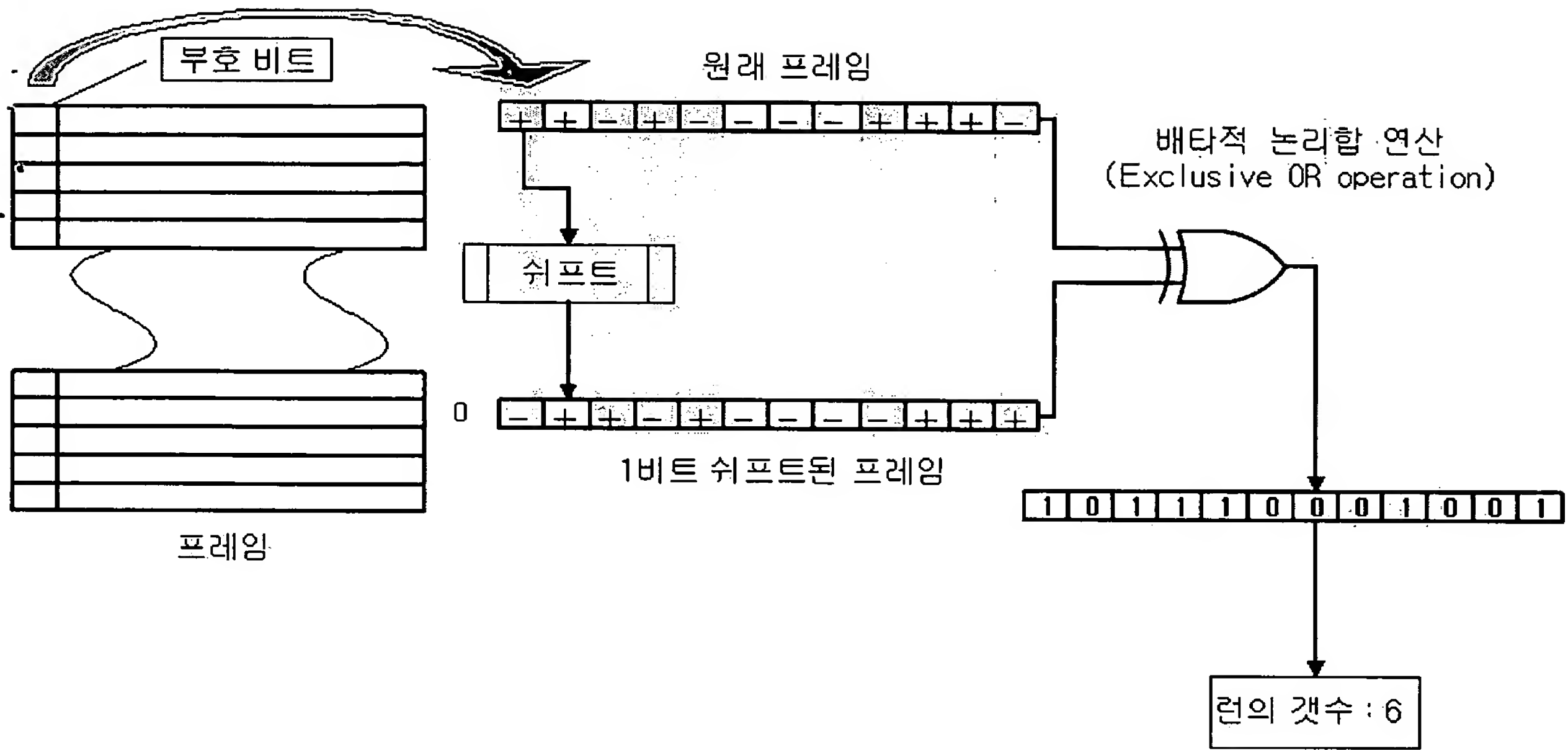
【도 4】



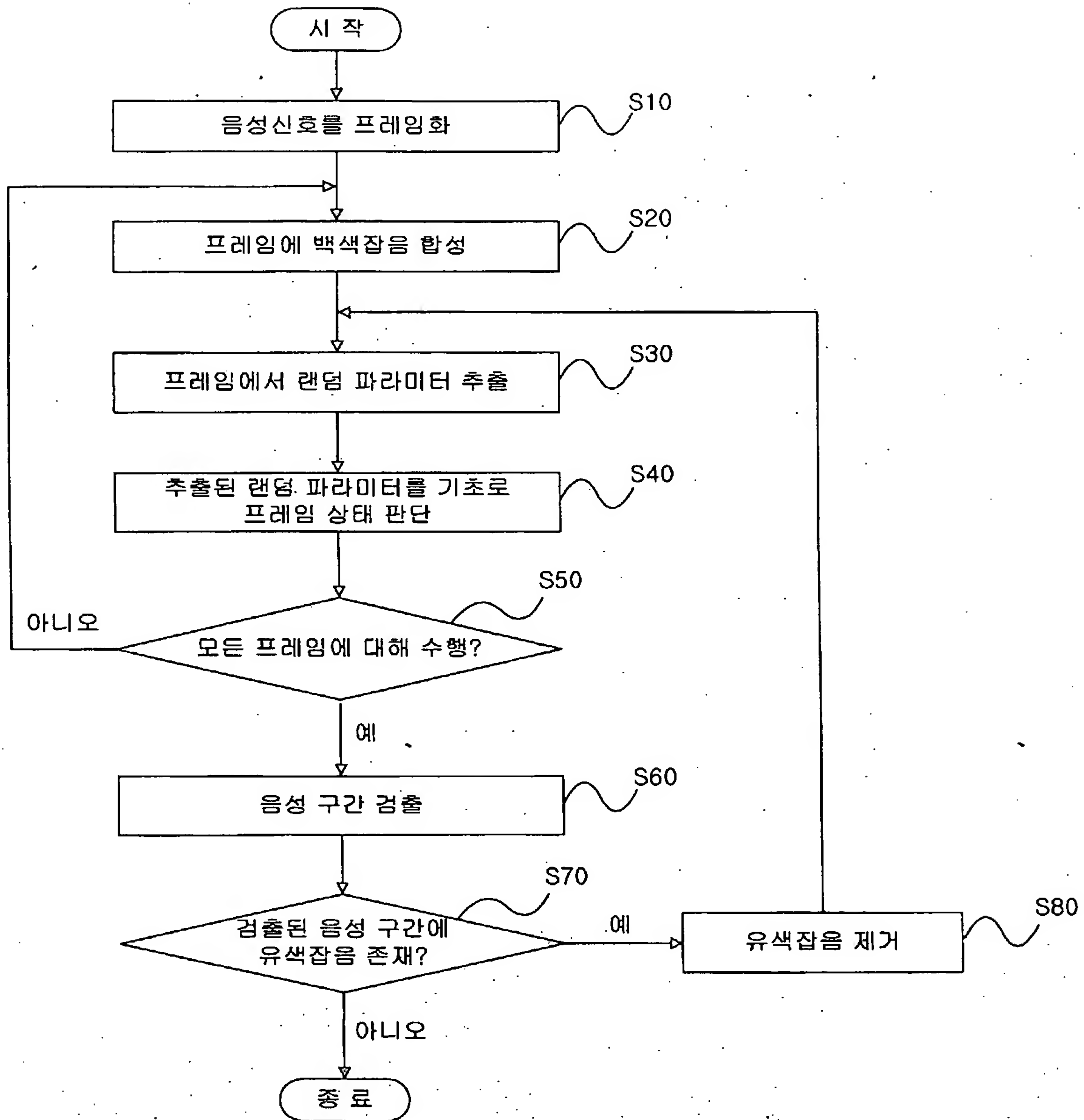
【도 5】



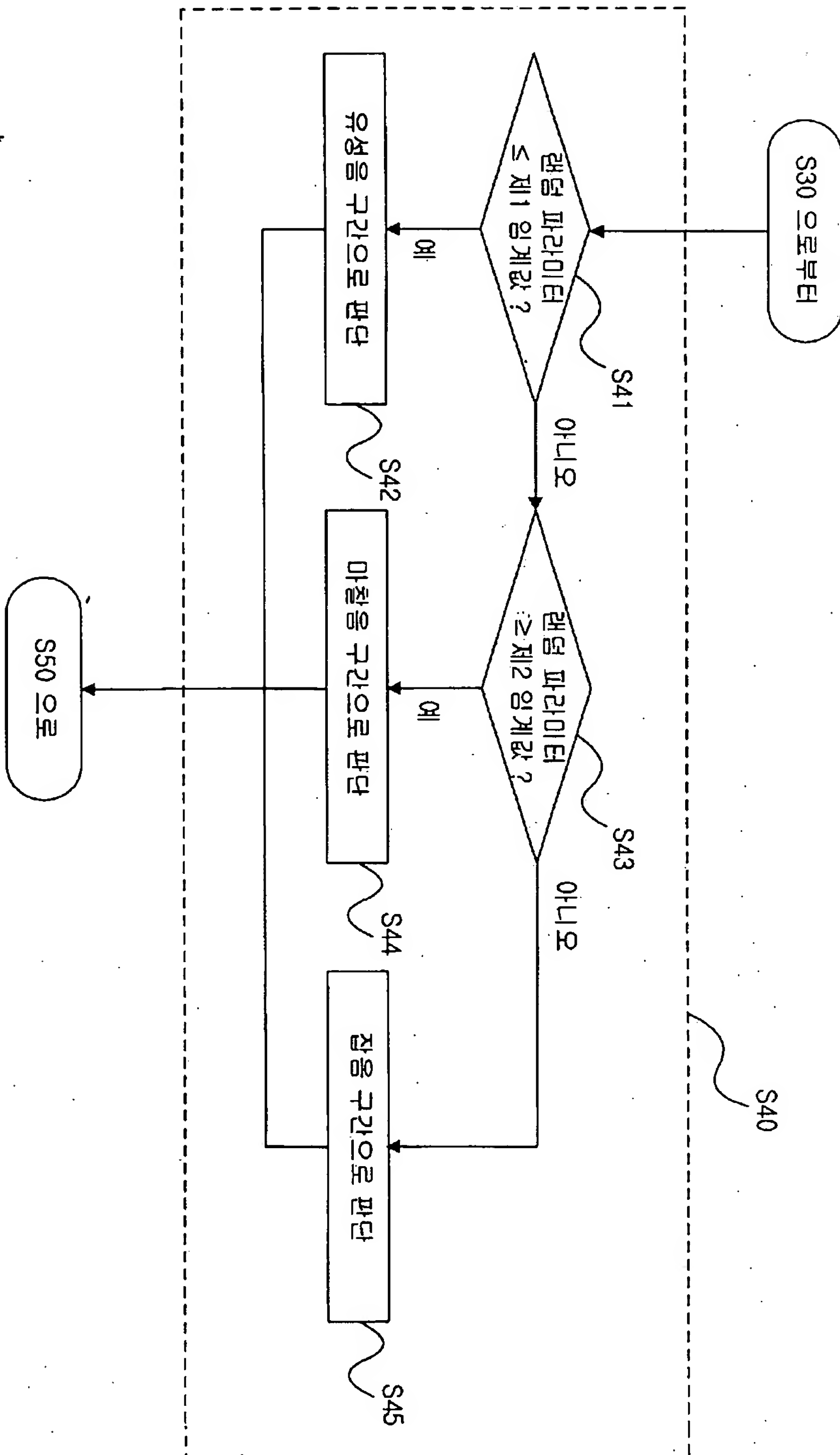
【도 6】



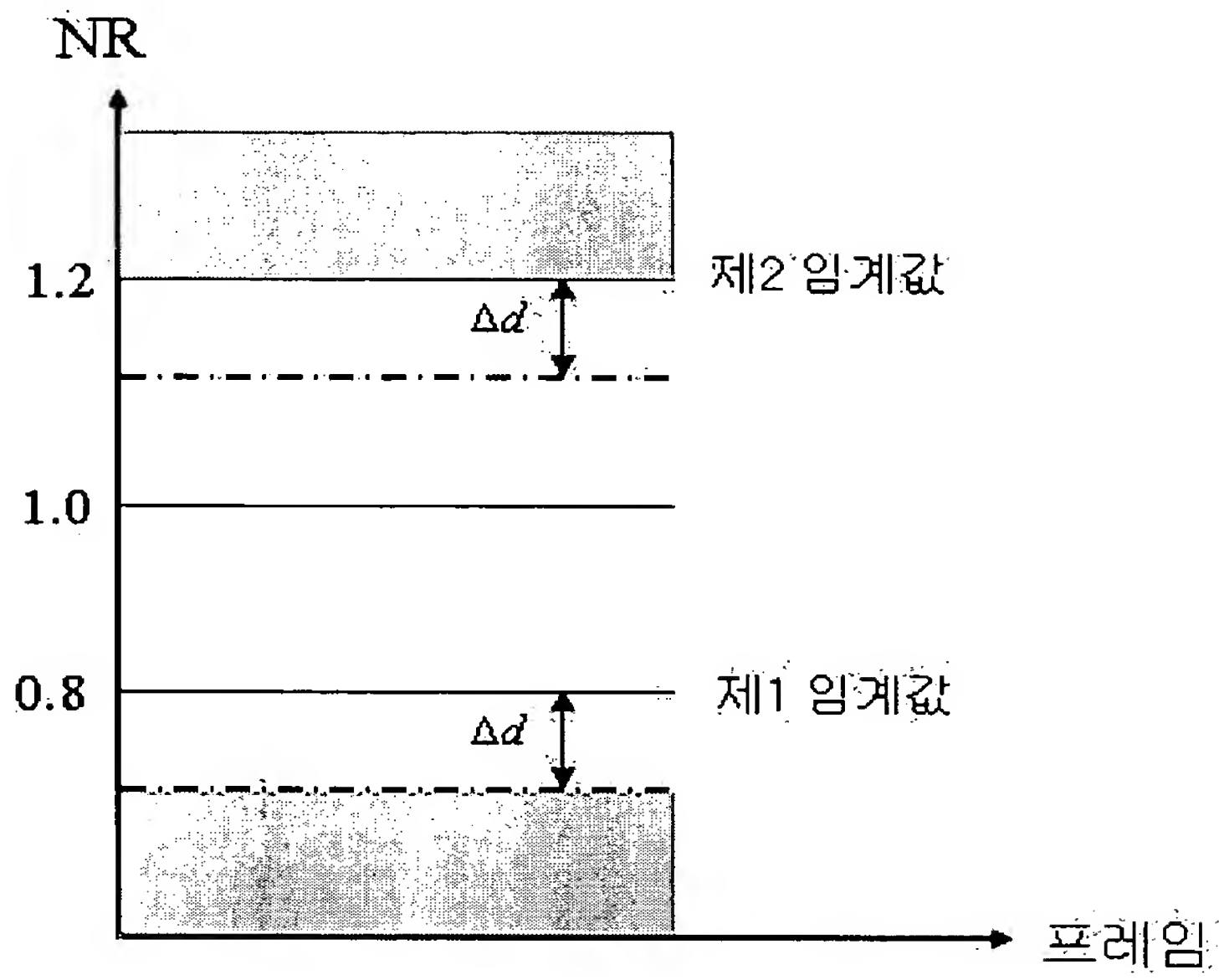
【도 7】



【도 8】

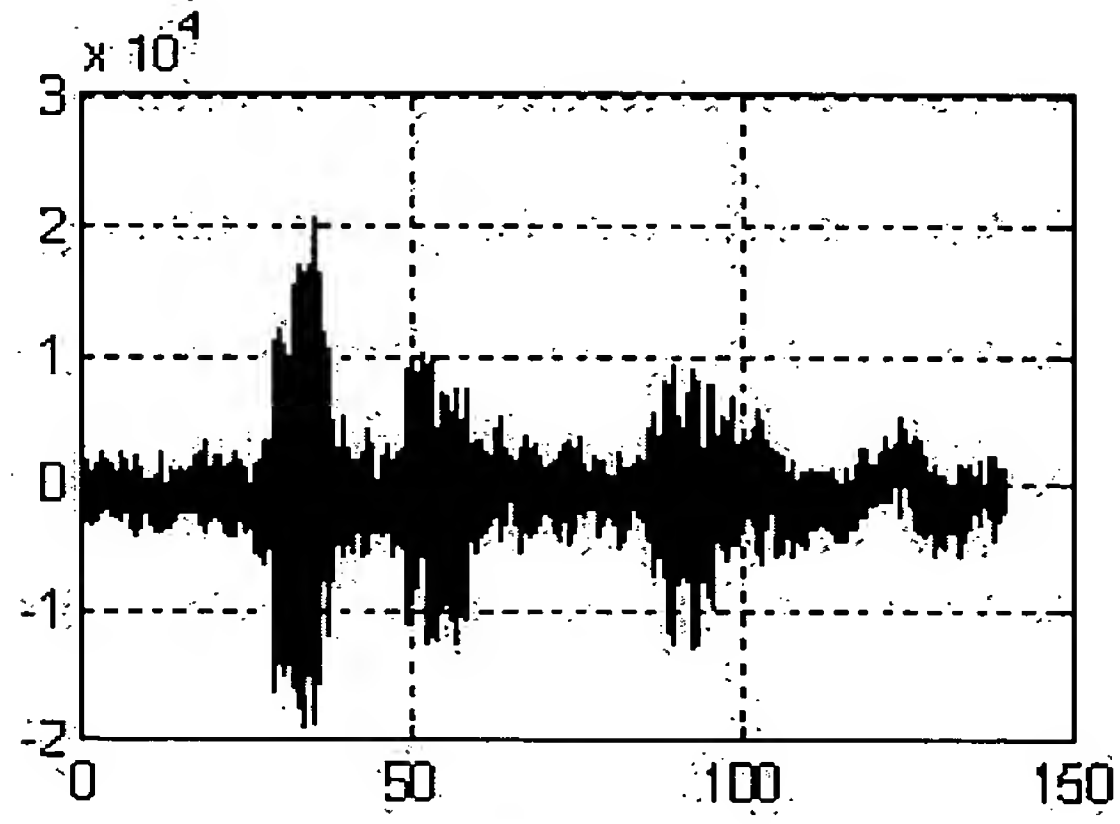


【도 9】

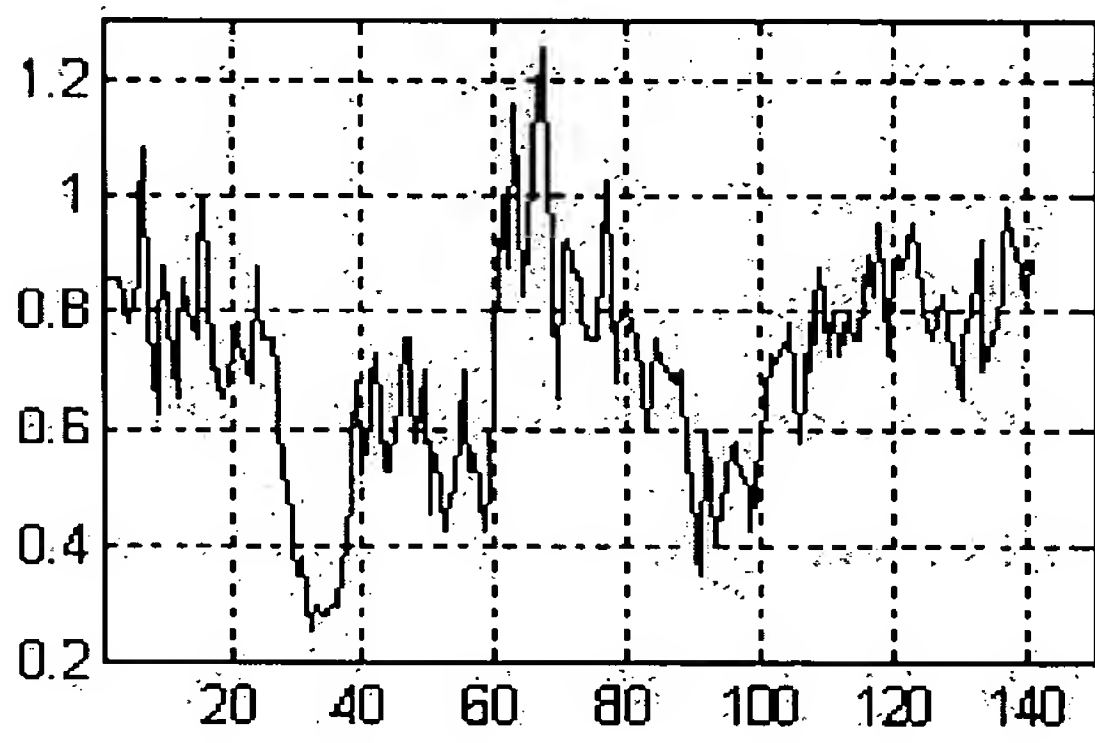


【도 10】

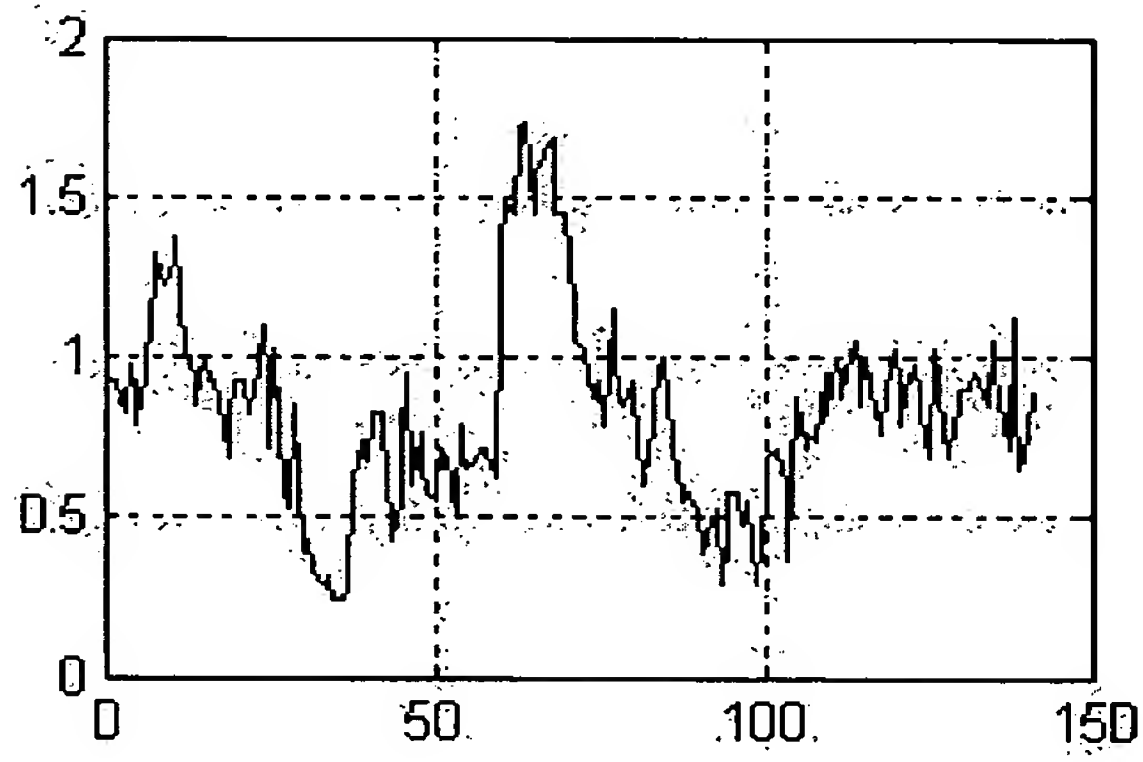
(a)



(b)

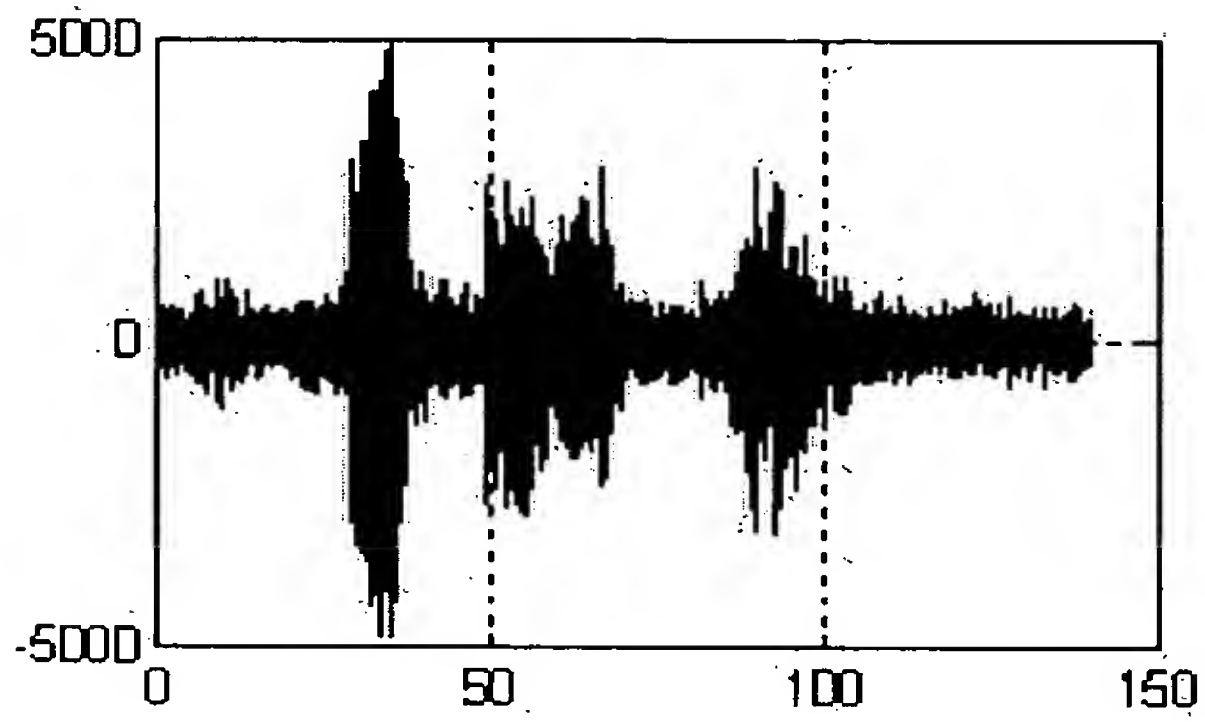


(c)

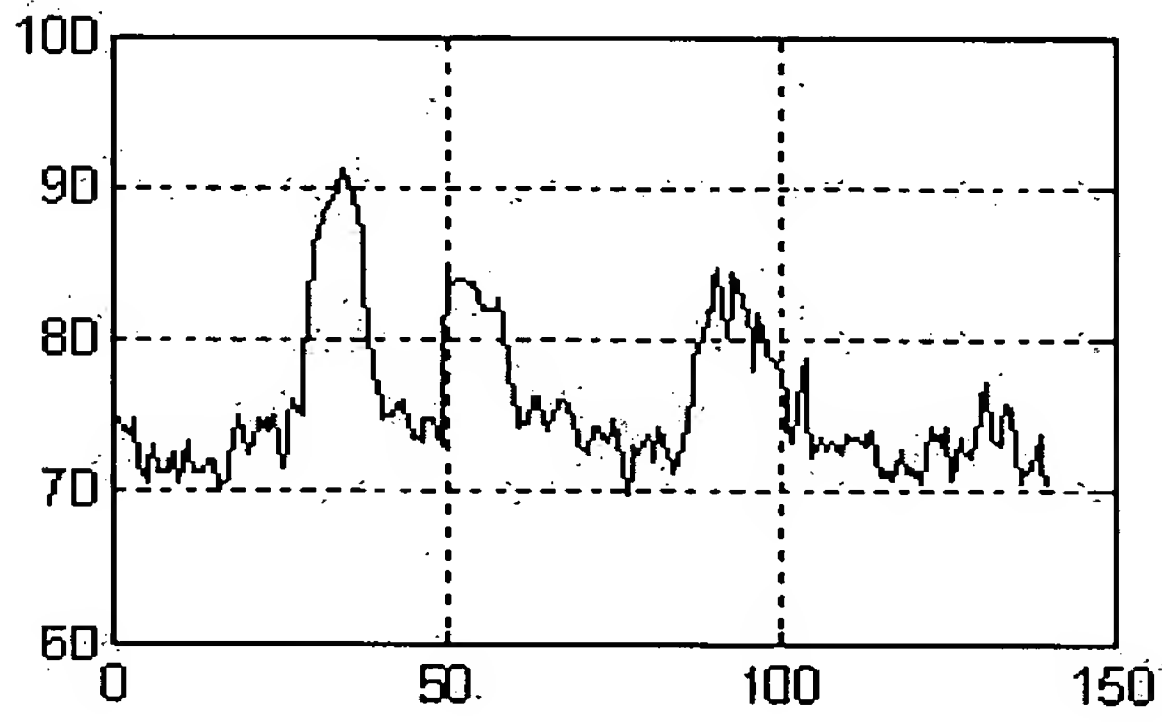


【도 11】

(a)



(b)



(c)

